Perspectives on the Theory and Practice of Belief Functions

Glenn Shafer
School of Business, University of Kansas, Lawrence, Kansas

ABSTRACT

The theory of belief functions is a generalization of the Bayesian theory of subjective probability judgment. The author's 1976 book, A Mathematical Theory of Evidence, is still a standard reference for this theory, but it is concerned primarily with mathematical foundations. Since 1976, considerable work has been done on interpretation and implementation of the theory. This article reviews this work, as well as newer work on mathematical foundations. It also considers the place of belief functions within the broader topic of probability and the place of probability within the larger set of formalisms used by artificial intelligence.

KEYWORDS: ***Bayesian theory, belief functions, Dempster–Shafer theory, independence, inner measures, interactive systems, join trees, lower probabilities, multivalued mappings, probability propagation, random sets, statistical inference***

1. INTRODUCTION

The theory of belief functions provides one way to use mathematical probability in subjective judgment. It is a generalization of the Bayesian theory of subjective probability. When we use the Bayesian theory to quantify judgments about a question, we must assign probabilities to the possible answers to that question. The theory of belief functions is more flexible; it allows us to derive degrees of belief for a question from probabilities for a related question. These degrees of belief may or may not have the mathematical properties of probabil-

Address correspondence to Glenn Shafer, School of Business, Summerfield Hall, University of Kansas, Lawrence, KS 66045.

International Journal of Approximate Reasoning 1990; 4:323–362

655 Avenue of the Americas, New York, NY 10010 0888-613X/90/$3.50

ities; how much they differ from probabilities will depend on how closely the two questions are related.

Examples of what we would now call belief function reasoning can be found in the late seventeenth and early eighteenth centuries, well before Bayesian ideas were developed. In 1689, George Hooper gave rules for combining testimony that can be recognized as special cases of Dempster's rule for combining belief functions (Shafer [90]). Similar rules were formulated by Jakob Bernoulli in his *Ars Conjectandi*, published posthumously in 1713, and by Johann-Heinrich Lambert in his *Neues Organon*, published in 1764 (Shafer [84]).

Examples of belief function reasoning can also be found in more recent work by authors who were unaware of the seventeenth and eighteenth century work. For example, Per Olof Ekelöf, a Swedish legal scholar, reinvented Hooper's rules in the early 1960s (Shafer [90], Gärdenfors et al. [37]).

In its present generality, however, the theory of belief functions is due to Arthur P. Dempster and myself. Dempster set out the basic ideas of the theory in a series of articles in the 1960s. I developed the theory further and established its terminology and notation in 1976 in *A Mathematical Theory of Evidence* [82]. Because of the role of Dempster and myself, the theory is sometimes called the "Dempster–Shafer theory."

My 1976 monograph is still the most comprehensive source of information on belief functions. Unfortunately, it says little about interpretation, implementation, or computation. It is concerned primarily with mathematical foundations.

During the past fifteen years, a great deal has been learned about the interpretation, implementation, and computation of belief functions, and fresh progress has been made on the mathematical foundations as well. This work is scattered in journals in a wide variety of fields, including statistics, psychology, philosophy, engineering, accounting, and artificial intelligence. No one has drawn the new work together in a comprehensive way.

This article is primarily a summary of my own current views. It pulls together some of the strands in the scattered literature, but it falls short of the comprehensive review that is needed. Though the bibliography is lengthy, it is not comprehensive. My failure to mention particular contributions should not be taken as an indication of disinterest or disagreement.

I begin with some opinions about the place of belief functions in the larger tool drawer of probabilistic methods and about the place of this whole tool drawer in artificial intelligence (Sections 2 and 3). These opinions are personal. I expect that few readers will agree with them fully. They may nonetheless serve to dispel some misunderstandings about the scope of the theory of belief functions.

After this general introduction, I introduce the basic ideas of belief functions, in a way that should be accessible to readers with no previous familiarity with the topic (Section 4), and I review the many competing mathematical formulations of the theory (Section 5).

I then turn to issues of interpretation and semantics. I explain that belief function degrees of belief should not be interpreted as lower bounds on unknown probabilities (Section 6). Instead, they should be given a process-oriented semantics; they are the result of deliberately drawing a certain analogy (Section 7).

I then turn to issues closer to implementation. I discuss how the concept of independence is used in belief function reasoning (Section 8) and how belief functions can and cannot be used in reasoning about frequency distributions (Section 9). I discuss how the computational complexity of belief functions can be dealt with (Section 10). And I discuss the extent to which belief function reasoning can be automated (Section 11).

Finally, I discuss briefly a number of other topics in belief function theory—generalizations, decision methods, consensus, infinite frames, weights of evidence, and other mathematical advances (Section 12).

2. THE PLACE OF BELIEF FUNCTIONS IN PROBABILITY

In my earliest work on belief functions [81–83] I took a youthfully ambitious view of their possible role. I saw belief functions as a very general tool for subjective judgment. Almost every item of evidence could be represented, I suggested, by a belief function.

By the early 1980s, however, I was expressing a more sober view (see especially Ref. 86). Then, as now, I saw belief functions as one among many tools for making subjective judgments using probability. Other such tools include Bayesian probability, Fisherian tests of significance, and Neyman–Pearson confidence intervals.

It was once customary to equate Bayesian methods with subjective probability and to identify Fisherian and Neyman–Pearson methods with a frequentist conception of probability. These were the terms of the debate that raged between Bayesians and frequentists in the 1960s. In recent years, however, the debate has cooled, and both sides have acknowledged that subjectivity is involved in all applications of probability to real problems.

In recent work [96], I have tried to give a foundation to this emerging consensus by advancing the thesis that the mathematical theory of probability is really the theory of an ideal picture in which belief, fair price, and knowledge of the long run are bound together. Probabilities in this ideal picture are long-run frequencies, but they are also degrees of belief, because they are known and nothing else that is relevant is known. This ideal picture seldom occurs in nature, but there are many ways to make it relevant to real problems. In some cases, the ideal picture serves as a standard of comparison—Fisherian tests of significance are often used to compare correlations in real problems to accidental correlations in the ideal picture. In other cases, we simulate the ideal picture (we obtain random numbers from tables or numerical algorithms) and then deliberately entangle this simulation in a real problem (we use the random numbers to randomize experiments or draw random samples). In other cases, we draw an analogy between the state of knowledge in the ideal picture and our evidence in a real problem.

Subjectivity is involved in all these applications of probability, not only because it is involved in the ideal picture, but also because the comparison, entanglement, or analogy that relates the ideal picture to the real problem requires subjective judgment.

Bayesian methods rely on one particular type of analogy that can be drawn to the ideal picture. In this analogy, we compare our evidence about the answer to a question of interest to knowledge of frequencies in the ideal picture. This analogy is strongest when extensive frequency data are available for our problem (in this case, it is customary to talk about "empirical Bayes"). When our evidence does not consist of frequencies, the analogy may or may not be convincing.

Belief functions arise when the Bayesian type of analogy is drawn between the ideal picture and one question, and then the resulting probabilities are examined for their relevance to a related question that is of greater interest to us but for which we do not have a convincing direct Bayesian analogy. Here, as in the Bayesian case, there is no guarantee that the argument by analogy will be convincing. But if there are a number of questions that are related to the question of interest, then we may have a chance to find a successful analogy between the ideal picture and our evidence for at least one of these questions (Shafer and Tversky [101]).

Which of these many different ways of using probability is most important in practice? Fisherian and Neyman–Pearson applications are by far the most common and most important. Bayesian applications have been steadily increasing in recent decades. There are still few belief function applications. I expect both the dominance of Fisherian and Neyman–Pearson applications and the steady increase in Bayesian applications to continue. I am also optimistic that the number of belief function applications will continue to increase as the ideas of the theory become more widely understood.

3. THE ROLE OF PROBABILITY IN ARTIFICIAL INTELLIGENCE

What role has probability played in AI, and what role should it play? Does probability provide a general theory for handling uncertainty in AI? In this section, I review the history of probability in AI, and I argue that we should be modest and realistic about the future prospects for probability in AI.

Historically, probability has not played a strong role in the theory of AI. Beginning with the inception of AI as a field in the 1950s, the theory of AI has been dominated by logic. On the other hand, there has always been a role for probability in AI applications. Standard probabilistic and statistical methods have long been used in AI work in pattern recognition and learning (Nilsson [71], Duda and Hart [31]), and recently probabilistic ideas and techniques have been used in expert systems (Spiegelhalter [112]).

The drive to make logic a foundation for AI has faltered in recent years. At the beginning of the 1980s, logicians in AI were still optimistic about extending logic to deal with commonsense or uncertain reasoning, and the term "nonmonotonic logic" was coined as a general name for such extensions. As the 1980s drew to a close, however, the promise of nonmonotonic logic faded, and some of its most prominent original proponents questioned the fundamental idea of using logic for practical reasoning (see especially McDermott [66]).

Does the faltering role of logic mean that it should be replaced by probability as a general foundation for AI? A number of probabilists, including Cheeseman [12] and Pearl [74], believe that it should. These probabilists point out that uncertainty is involved in the vast majority of AI problems, and they repeat general arguments by Savage [79], Cox [15], and others to the effect that probability is always appropriate for dealing with uncertainty. In my opinion, this line of thinking is misguided. Common sense tells us that there are many situations involving uncertainty in which the theory of probability is not useful. The arguments advanced by Savage and Cox should be seen as clarifications of subjective probability, not as demonstrations that it is always appropriate (Shafer [91, 95]). The cause of probability in AI will be better served, in the long run, by a realistic and modest assessment of its limitations and potential.

The mere fact that there is uncertainty in a problem does not mean that the theory of probability is useful in solving the problem. As I pointed out in the preceding section, the theory of probability is really about a special ideal picture, not about all situations involving uncertainty. Bringing the theory of probability to bear on a particular problem means relating it in some way to the ideal picture. This requires imagination, and success is not guaranteed. Even if there is a way to bring the ideal picture to bear on a problem, doing so may not be worth the trouble. In many AI applications, it is more sensible to generate one plausible guess or plan than to weigh probabilities. In others, it is more efficient to give more reliable sources of information priority in a hierarchical system of control than to weigh the probabilities of each report from these sources (Brooks [11], Cohen [14]).

Since the ideal picture of probability involves frequencies, probability is most easily applied when relevant frequencies are available. Artificial intelligence is no exception to the rule that frequentist applications far outnumber subjective applications. The applications to pattern recognition and learning are predominantly frequentist, as are the most successful current applications of probability to expert systems. Pearl [74] has emphasized the potential of subjective judgement in causal models, arguing that causal models are persuasive even when frequencies are replaced by subjective guesses. But this persuasiveness is not always matched by reliability. Numerous studies have shown that simple actuarial rules of thumb, when they are available, are more reliable than subjective clinical judgment (Dawes et al. [16]).

Finding ways of using probability—or alternatives to probability—when directly relevant frequencies are not available should be seen as a difficult task. It is not a task that can be disposed of by general normative arguments such as those of Savage. It is not a task that can be disposed of by finding the right alternative calculus, such as the theory of belief functions or Zadeh's theory of possibility measures (Zadeh [136]). It is a task that must be dealt with in the context of each application.

4. THE BASIC IDEAS OF BELIEF FUNCTIONS

The theory of belief functions is based on two ideas: the idea of obtaining degrees of belief for one question from subjective probabilities for a related question, and Dempster's rule for combining such degrees of belief when they are based on independent items of evidence.

The simplest way to illustrate these ideas is to go back to the topic addressed by George Hooper in 1689, the reliability of testimony. The belief function approach to testimony is to derive degrees of belief for statements made by witnesses from subjective probabilities for the reliability of these witnesses.

Suppose that Betty tells me a tree limb fell on my car. My subjective probability that Betty is reliable is 0.9; my subjective probability that she is unreliable is 0.1. Since they are probabilities, these numbers add to 1. But Betty's statement, which must true if she is reliable, is not necessarily false if she is unreliable. So I say that her testimony alone justifies a 0.9 degree of belief that a limb fell on my car, but only a zero degree of belief (not a 0.1 degree of belief) that no limb fell on my car. This zero does not mean that I am sure that no limb fell on my car, as a zero probability would; it merely means that Betty's testimony gives me no reason to believe that no limb fell on my car. The 0.9 and the zero together constitute a belief function.

Dempster's rule is based on the standard idea of probabilistic independence, applied to the questions for which we have subjective probabilities. I can use the rule to combine evidence from two witnesses if I consider the first witness's reliability subjectively independent (before I take account of what the witnesses say) of the second's reliability. (This means that finding out whether one witness is reliable would not change my subjective probability for whether the other is reliable.) The rule uses this subjective independence to determine joint probabilities for the various possibilities as to which of the two are reliable.

After using independence to compute joint probabilities for who is reliable, I must check whether some possibilities are ruled out by what the witnesses say. (If Betty says a tree limb fell on my car and Sally says nothing fell on my car, then they cannot both be reliable.) If so, I normalize the probabilities of the remaining possibilities so they add to 1. This is an example of probabilistic conditioning, and it may destroy the initial independence. (After I notice that Betty and Sally have contradicted each other, their reliabilities are no longer subjectively independent for me. Now finding out that one is reliable would tell me that the other is not.) After the normalization, I determine what each possibility for the reliabilities implies about the truth of what the witnesses said, and I use the normalized probabilities to get new degrees of belief.

The net effect of Dempster's rule is that concordant items of evidence reinforce each other, conflicting items of evidence erode each other, and a chain of reasoning is weaker than its weakest link. To illustrate this, consider two independent witnesses, say Betty and Sally. Suppose the reliabilities of Betty and Sally are p_1 and p_2, respectively; Betty's testimony gives us a degree of belief p_1 in what she says and degree of belief 0 in its denial, while Sally's testimony gives us a degree of belief p_2 in what she says and degree of belief 0 in its denial. Then we can derive the following formulas:

- If Betty and Sally say exactly the same thing, our degree of belief in what they say will be $1-(1-p_1)(1-p_2)$.
- If they make different but consistent assertions, our degree of belief in both assertions being true will be p_1p_2.
- If they make contradictory assertions, our degree of belief in Betty's assertion will be $p_1(1-p_2)/(1-p_1p_2)$, and our degree of belief in Sally's assertion will be $p_2(1-p_1)/(1-p_1p_2)$.

These formulas are derived in Chapter 4 of Ref. 82. They represent only the simplest examples of Dempster's rule. When we combine more complex belief functions, Dempster's rule becomes too complex to be represented informatively through simple formulas.

5. THE MATHEMATICAL FORMALISM OF BELIEF FUNCTIONS

The basic ideas of belief functions can be formalized mathematically in a variety of ways. In my 1976 monograph, I defined belief functions axiomatically, and I defined Dempster's rule by a formula. Other approaches include multivalued mappings, compatibility relations, random subsets, and inner measures. This section is concerned with these alternative approaches and with the relations among them.

The choice among these alternatives should be seen as a matter of convenience in mathematical exposition and investigation. Since the alternatives are mathematically equivalent, it makes no fundamental difference which we take as the starting point in such mathematical work.

In applications, the starting point does make a great deal of difference. But for applications we need more than a mathematical definition or a set of axioms as a starting point. We need a metaphor that can serve as a guide in relating a practical problem to the theory and as a guide in assessing numbers to represent the strength of evidence in the practical problem. I will consider the problem of providing such metaphors in Section 7. The alternatives I discuss in this section are not such metaphors; they are merely mathematical formulations.

In the simple example of Betty's testimony, we started with two questions:

Q_1: Is Betty reliable?

Q_2: Did a tree limb fall on my car?

We had probabilities for Q_1, and we derived degrees of belief for Q_2. This process required no formal notation, because both Q_1 and Q_2 had only two possible answers: yes and no. In more complex examples, we will have questions Q_1 and Q_2 with many possible answers. To talk about such examples in general, we need a notation for each question's set of possible answers, a notation for the probabilities for Q_1 and the degrees of belief for Q_2, and a way of representing the constraints that an answer to Q_1 may put on the answer to Q_2.

We assume that each question comes with an exhaustive list of mutually exclusive answers. We know that exactly one of these answers is correct, but we do not know which one. We call such a set of answers a *frame*. Let S be the frame for Q_1, the question for which we have probabilities, and let T be the frame for Q_2, the question of interest.

Let us write $P(s)$ for the probability of the element s of S. Given these probabilities, and given a subset A of T, we want to derive a degree of belief $\mathrm{Bel}(A)$, our degree of belief that A contains the correct answer to Q_2.

An answer s to Q_1 may rule out a whole set of answers to Q_2. If A is a set of answers to Q_2, and s rules out all the answers in A's complement, $T - A$, then s tells us that the answer to Q_2 is somewhere in A. Thus the probability $P(s)$ should contribute to our belief in A. Our total degree of belief in A, $\mathrm{Bel}(A)$, should be the total probability for all answers s that rule out all answers in $T - A$. How shall we put this into symbols?

Here is where paths diverge. Multivalued mappings, compatibility relations, random subsets, and inner measures all provide different ways of specifying mathematically the answers to Q_2 ruled out by an answer to Q_1 and hence different ways of explaining the relation between the probabilities on S and the degrees of belief on T.

Multivalued Mappings

Let us first consider multivalued mappings, which were used by Dempster in his early articles (e.g., Dempster [18]).

Let us write $\Gamma(s)$ for the subset of T consisting of the answers to Q_2 that are *not* ruled out by s. In this notation, s tells us that the answer to Q_2 is somewhere in A whenever

$$\Gamma(s) \subseteq A$$

The degree of belief Bel(A) will be the total probability for all answers s that satisfy this condition. In symbols,

$$\text{Bel}(A) = P\{s|\Gamma(s) \subseteq A\} \tag{1}$$

The mapping Γ is a *multivalued mapping* from S to T. Formally, a *belief function* is any function Bel given by (1) for some multivalued mapping Γ and some probability measure P.

Now consider two belief functions Bel_1 and Bel_2 on T, which we judge to be based on independent items of evidence. Each belief function will be based on its own probability space and its own multivalued mapping from that probability space to T. We may write S_1 and S_2 for the two probability spaces, P_1 and P_2 for the two probability measures, and Γ_1 and Γ_2 for the two multivalued mappings. Dempster's rule is a rule for combining Bel_1 and Bel_2 to obtain a belief function Bel on T that represents the pooling of two items of evidence. How do we describe Dempster's rule in terms of S_1, S_2, P_1, P_2, Γ_1, and Γ_2?

We can answer this question by using S_1, S_2, P_1, P_2, Γ_1, and Γ_2 to construct a probability space S, a probability measure P, and a multivalued mapping Γ from S to T. The belief function Bel given by combining Bel_1 and Bel_2 by Dempster's rule will be the belief function given by (1) using S, P, and Γ.

Independence of the two items of evidence means that we can make initial joint probability judgments about the two questions answered by S_1 and S_2 by forming the product measure $P_1 \times P_2$ on $S_1 \times S_2$. It also means that what an element s_1 of S_1 tells us about Q_2 does not affect what an element s_2 of S_2 tells us about Q_2; s_1 and s_2 together tell us only that the answer to Q_2 is in the intersection $\Gamma_1(s_1) \cap \Gamma_2(s_2)$. If this intersection is empty for some (s_1, s_2), then s_1 and s_2 are telling us contradictory things about Q_2, and one of them must be wrong. So we must condition the product measure $P_1 \times P_2$ on the set of (s_1, s_2) for which $\Gamma_1(s_1) \cap \Gamma_2(s_2)$ is not empty. We let S be the subset of $S_1 \times S_2$ consisting of (s_1, s_2) for which $\Gamma_1(s_1) \cap \Gamma_2(s_2)$ is not empty, and we let P be the probability measure on S obtained by conditioning $P_1 \times P_2$ on S. Finally, we let Γ be the multivalued mapping from S to T given by

$$\Gamma(s_1, s_2) = \Gamma_1(s_1) \cap \Gamma_2(s_2)$$

This completes the construction of S, P, Γ and hence the statement of Dempster's rule.

To summarize verbally: Dempster's rule says to form the product probability space, condition it by eliminating pairs that map to disjoint subsets of T, and then obtain a belief function by mapping each remaining pair to the intersection of the subsets to which the two elements of the pair are mapped.

Compatibility Relations

The multivalued mapping Γ from S to T tells us, for each element s of S, which elements of T are possible answers to Q_2 if s is the correct answer to Q_1. It tells us, in other words, which s's are compatible with which t's. This information can also be represented by specifying the set C of all ordered pairs (s, t) such that s is compatible with t. This is a "relation"—a subset of the Cartesian product $S \times T$. The relation C is related to the multivalued mapping Γ by

$$C = \{(s, t) | t \in \Gamma(s)\}$$

and

$$\Gamma(s) = \{t | (s, t) \in C\}$$

In terms of C, (1) becomes

$$\mathrm{Bel}(A) = P\{s | \{t | (s, t) \in C\} \subseteq A\}$$

or

$$\mathrm{Bel}(A) = P\{s | \text{if } (s, t) \in C, \text{ then } t \in A\} \tag{2}$$

In Ref. 94 I call C a *compatibility relation* and develop the mathematics of belief functions from (2). This is also the approach taken by Lowrance [64] and by Shafer and Srivastava [100].

In truth, the choice between multivalued mappings and compatibility relations is scarcely more than a choice of terms. Many treatments of mathematical set theory (e.g., Kelley [49]) define mappings as relations. Moreover, Dempster used the word "compatible" repeatedly in explaining the meaning of his multivalued mappings.

Random Subsets

We have assigned a subset of T, $\Gamma(s)$, to each s, and we have assigned a probability $P(s)$ to each s. If we think of the probability as being attached to the subset instead of to s, then we have, in effect, defined a random subset of T.

The subset $\Gamma(s)$ might be the same for different s's. Therefore, to find the total probability that the random subset will be equal to B, we must add the probabilities of all the s's for which $\Gamma(s)$ is equal to B. The degree of belief, $\mathrm{Bel}(A)$, is the total probability that the random subset is contained in A.

In this setting, Dempster's rule is a rule for combining two random subsets to obtain a third, or, more precisely, a rule for combining two probability distributions for random subsets to obtain a third. We assume that the random subsets are probabilistically independent, we intersect them, and then we condition the probability distribution for the intersection on its being nonempty.

This approach to the mathematics of belief functions was emphasized by Nguyen [70], Goodman and Nguyen [40], and Shafer et al. [102]. It is convenient for advanced mathematical exposition, because the idea of a random subset is well established among mathematical probabilists (Matheron [67]).

The Axiomatic Approach

Another approach is to characterize belief functions directly in terms of their mathematical properties. We simply list a set of axioms that a belief function Bel must satisfy. And we use a formula to define Dempster's rule for combining two belief functions Bel_1 and Bel_2. This was my approach in early papers [81–83, 85]. It is related to earlier mathematical work by Choquet [13].

In my 1976 monograph, I also gave a more transparent characterization for the case where T is finite. We assign a non-negative number $m(B)$ to each subset B of T. Intuitively, $m(B)$ is the probability that the random subset is equal to B. We require that $m(\varnothing) = 0$, where $\varnothing$ is the empty set, and that the $m(B)$'s add to 1. The function m is called the *basic probability assignment*. We define the function Bel by

$$\mathrm{Bel(A)} = \sum \{\mathrm{m(B)} | \mathrm{B} \subseteq \mathrm{A}\} \tag{3}$$

As it turns out, the $m(B)$ can then be recovered from the Bel(A) by the formula

$$m(B) = \sum \{(-1)^{|A|} \mathrm{Bel}(A) | A \subseteq B\} \tag{4}$$

The functions m and Bel are said to be Möbius transforms of each other (Rota [77]). Dempster's rule for combining two belief functions Bel_1 and Bel_2 can be defined by a relatively simple rule in terms of the corresponding basic probability assignments m_1 and m_2; we define $m(B)$ by

$$m(B) = \frac{\sum \{m_1(B_1)m_2(B_2) | B_1 \cap B_2 = B\}}{\sum \{m_1(B_1)m_2(B_2) | B_1 \cap B_2 \neq \varnothing\}} \tag{5}$$

The belief function Bel resulting from the combination can then be obtained from m by using (3).

Many other authors, including Gordon and Shortliffe [41, 42], have used basic probability assignments to explain the theory of belief functions.

Inner Probability

The idea of deriving minimal degrees of belief for some sets from probabilities for other sets has long been familiar in abstract probability theory in the context of "inner measures" (Halmos [43]) or "inner probabilities" (Neveu

[69]). With attention to a few technicalities, we can relate belief functions to the idea of inner measure or inner probability.

It is easiest to explain this using the compatibility relation C that relates the frames S and T. The set C is itself a frame—it consists of the possible answers to the joint question formed by compounding the question answered by S and the question answered by T. A Bayesian approach to our problem would construct a probability measure on C. The belief function approach stops short of completing this construction. We stop when we have constructed a probability measure P on S. We then extend P to an inner probability on C.

Let us review the definition of inner probability in the finite case. Recall that an algebra of subsets of a set X is a collection of subsets that includes X and the empty set $\varnothing$ and also includes $A \cap B$, $A \cup B$, $X - A$, and $X - B$ whenever it includes A and B. Given a probability measure Q defined only an algebra $\mathcal{A}$ of subsets of a finite set X, the inner probability of Q is the function Q_* defined by

$$Q_*(A) = \max\{Q(B) | B \in \mathcal{A} \text{ and } B \subseteq A\} \tag{6}$$

for every subset A of X. Intuitively, $Q_*(A)$ is the degree to which the probabilities for the elements of $\mathcal{A}$ force us to believe A.

Let $\mathcal{A}$ denote the collection of all subsets of C of the form $C \cap (R \times T)$, where R is a subset of S. This is an algebra of subsets of C. Since the subset $C \cap (R \times T)$ of C has the same meaning (*qua* assertion about the answer to Q_1) as the subset R of S, it is natural to define a probability measure Q on $\mathcal{A}$ by setting $Q(C \cap (R \times T)) = P(R)$. With this definition, (6) becomes

$$Q_*(A) = \max\{P(R) | R \subseteq S \text{ and } C \cap (R \times T) \subseteq A\} \tag{7}$$

for every subset A of C.

What belief should we give to a subset U of T? It is natural to answer by looking at the value of (7) for the subset of C that corresponds to U, namely $C \cap (S \times U)$. This is

$$\begin{aligned} &Q_*(C \cap (S \times U)) \\ &\quad = \max\{P(R) | R \subseteq S \text{ and } C \cap (R \times T) \subseteq C \cap (S \times U)\} \\ &\quad = \max\{P(R) | R \subseteq S \text{ and if } s \in R \text{ and } (s, t) \in C, \text{ then } t \in U\} \\ &\quad = P\{s | \text{if } (s, t) \in C, \text{ then } t \in U\} \end{aligned}$$

which is the same as formula (2) for $\mathrm{Bel}(U)$.

Thus a belief function is simply the inner measure of a probability measure—or, more precisely, the restriction to a subalgebra of an inner measure obtained from a probability measure on a possibly different subalgebra.

This connection between inner measures and belief functions must have been known for some time to many students of belief functions. To the best of my knowledge, however, it has appeared in the literature only in the past few years. The only references I know are Ruspini [78] and Fagin and Halpern [33].

6. BELIEF FUNCTION DEGREES OF BELIEF ARE NOT LOWER BOUNDS

In this section, I will review the point, well established in the literature on belief functions, that belief function degrees of belief should not be interpreted as bounds on unknown true probabilities. Such an interpretation seems plausible when we consider only a single belief function, but it breaks down when we consider belief functions that represent different and possibly conflicting items of evidence. Most important, a probability-bound interpretation is incompatible with Dempster's rule for combining belief functions. If we make up numbers by thinking of them as lower bounds on true probabilities, and we then combine these numbers by Dempster's rule, we are likely to obtain erroneous and misleading results.

In order to see how the degrees of belief given by a belief function might be thought of as lower bounds on probabilities, consider again my 0.9 belief that a limb fell on my car and my zero belief that no limb fell on my car. These degrees of belief were derived from my 0.9 and 0.1 subjective probabilities for Betty being reliable or unreliable. Suppose these subjective probabilities were based on my knowledge of the frequency with which witnesses like Betty are reliable. Then I might think that the 10% of witnesses like Betty who are not reliable make true statements a definite (though unknown) proportion of the time and false statements the rest of the time. Were this the case, I could think in terms of a large population of statements made by witnesses like Betty. In this population, 90% of the statements would be true statements by reliable witnesses, x% would be true statements by unreliable witnesses, and $(10-x)$% would be false statements by unreliable witnesses, where x is an unknown number between 0 and 10. The total chance of getting a true statement from this population would be $(90+x)$%, and the total chance of getting a false statement would be $(10-x)$%. My degrees of belief of 0.9 and zero are lower bounds on these chances; since x is anything between 0 and 10, 0.9 is the lower bound for $(90+x)$%, and zero is the lower bound for $(10-x)$%.

As this example suggests, a *single* belief function is always a consistent system of probability bounds. For any belief function Bel over any finite frame T, there will exist a class of probability distributions $\mathcal{P}$ such that

$$\mathrm{Bel}(A) = \min_{P \in \mathcal{P}} P(A) \tag{8}$$

for every subset A of T. [There are many ways of seeing that this is true. One way is to recall that Bel(A) is the sum of $m(B)$ for all B contained in A. Consider the different probability distributions obtained by distributing the mass $m(B)$, for each B, among the elements of B. If B is contained in A, then all the mass has to fall in A, but if B is not contained in A, then it is possible to distribute it all outside of A. Hence the minimum probability that one of these distributions can give A is Bel(A).]

However, the degrees of belief given by belief functions should *not* be interpreted as lower bounds on some unknown true probability. Belief functions are not, in general, concerned with a well-defined reference population or with learning about the frequencies in this population. And differences between belief functions do not, in general, reflect disagreements about unknown true probabilities. When Betty says a limb fell on may car, and Sally says nothing fell on my car, they are disagreeing about whether something fell on my car, not about the true probability of something having fallen on my car.

Were we to insist on a probability-bound interpretation of belief functions, then we would be interested only in groups of belief functions whose degrees of belief, when interpreted as probability bounds, can be satisfied simultaneously. But when belief functions are given their proper interpretation, it is of no particular significance whether there exist probabilities that simultaneously satisfy the bounds defined by a whole group of belief functions. Consider two cases that might arise when we use belief functions to represent contradictory evidence from Betty and Sally.

CASE 1 Before hearing their testimony, we think highly of the reliability of both Betty and Sally. We represent Betty's evidence by a belief function that gives a 95% degree of belief to a limb having fallen on my car, and we represent Sally's evidence by a belief function that gives a 95% degree of belief to nothing having fallen on my car. In this case, the two belief functions are contradictory as probability bounds; if the true probability of a limb having fallen on my car is greater than 95%, then the true probability of nothing having fallen on my car cannot also be greater than 95%.

CASE 2 Before hearing their testimony, we think that both Betty and Sally are fairly unreliable. So in both belief functions, we assign a 35% degree of belief rather than a 95% degree of belief. In this case, the two belief functions define consistent probability bounds; the true probability of a limb having fallen on my car and of nothing having fallen on my car can both be greater than 35%.

From the belief function point of view, there is no conceptual difference between these two cases. In both cases, we can combine the two belief functions by Dempster's rule. In both cases, there is conflict in the evidence being combined, and normalization is required.

It can be shown that if no normalization is required in the combination of a group of belief functions by Dempster's rule, then there do exist consistent probabilities that simultaneously bound all the belief functions being combined as well as the belief function that results from the combination. This may also happen when normalization is required, as in case 1, but we cannot count on this. In general, a probability-bound interpretation of belief functions is inconsistent with normalization (Zadeh [137]).

Probability bounds do provide another way to use the ideal picture of probability in subjective judgment. I have called this the *lower-probability* approach in order to distinguish it from the belief function approach (Shafer [86]). The lower-probability approach has been elaborated by Smith [110], Good [38], Suppes [116], Levi [62], Kofler et al. [50], Nilsson [72], Snow [111], and others. It does not, in general, use Dempster's rule.

When we use the lower-probability approach in a practical problem, we are drawing an analogy between actual evidence and knowledge of bounds on unknown true probabilities for the question of interest. Like the belief function approach, the lower-probability approach is sometimes but not always appropriate and useful. To claim it is always appropriate is to fall into the trap of assuming that unknown true probabilities always exist. In general, they exist only if a population and sampling scheme are well defined. An unknown true probability for the truth of Betty's statement, for example, exists only if the population of true and false statements of witnesses like Betty is well defined.

In some cases, a lower-probability analysis will come very close to a Bayesian analysis. Indeed, if the bounds we consider are fairly tight—fairly close to defining a single probability distribution—then they may correspond to bounds we would consider to see how slight changes in our subjective probability distribution would affect the outcome of a Bayesian analysis. Some authors call this "sensitivity analysis" (Fishburn [35], Isaacs [45]); others call it "robust Bayesian analysis" (Kadane [47]).

One contrast between the belief function and lower-probability approaches is in how they combine a given belief function Bel on a frame T with evidence that establishes that the answer is in a subset A of T. The belief function approach is to combine Bel by Dempster's rule with a new belief function that gives degree of belief 1 to A and degree of belief 0 to every proper subset of A. This generalizes Bayesian conditioning; I have called it *Dempster's rule of conditioning*. The lower-probability approach is to consider all the probability distributions bounded from below by the belief function, condition them all on A, and then take lower bounds over the resulting conditional probabilities. The lower-probability approach produces weaker degrees of belief than the belief function approach (Shafer [86], Kyburg [58]).

Though applying (8) to an arbitrary class of probability distributions $\mathcal{P}$ does not always produce a belief function, it does produce a belief function surprisingly often (Wasserman [124]). Moreover, the lower-probability approach to conditioning a belief function produces another belief function (Fagin and Halpern [34]). These results show that belief functions are of greater interest to the lower-probability approach than one might have expected, but they do reconcile the two approaches from the point of view of application and interpretation.

The point that belief functions are not lower bounds on probability measures is well accepted in the literature on belief functions. It is true that Dempster used the idea of bounds on unknown probabilities as a didactic tool in several of his articles on belief functions (which he then called "upper and lower probabilities") in the 1960s. The idea of probability bounds was never basic to Dempster's work, however; his work differed in this respect from the earlier work of Smith [110] and Good [38]. In my 1976 monograph, where I introduced both the name "belief function" and the notation now associated with belief functions, I explicitly disavowed the interpretation of belief function degrees of belief as lower bounds over classes of probabilities (Ref. 82, p. ix). In later articles, I have amplified, emphasized, and repeated this disavowal (Shafer [86, 94], Shafer and Srivastava [100]). Dempster and other proponents of belief functions have seconded the disavowal (Dempster [24], Ruspini [78], Smets [109]).

7. THE SEMANTICS OF BELIEF FUNCTIONS

As I noted in Section 5, the use of belief functions in practical problems requires metaphors that can guide us in relating practical problems to the theory and in assessing the strength of evidence numerically. This is the practical meaning of calls for a "semantics" for belief functions. In this section, I review the metaphors that I have suggested in the past as well as some of the suggestions others have made.

In my mind, the simplest and most effective metaphor for belief functions is the metaphor of the witness who may or may not be reliable. In many cases, the example of such a witness can serve as a standard of comparison for the strength of evidence. We can assess given evidence by saying that it is comparable in strength to the evidence of a witness who has a certain chance of being reliable.

A witness testifying to a specific proposition leads to a relatively simple belief function—one that gives a specified degree of belief to that proposition and its consequences and zero degree of belief to all other propositions. Arbitrarily complex belief functions can be built up by combining such simple belief functions (Shafer [82], p. 200), but in some cases we may want to produce complex belief functions more directly, in order to represent evidence that conveys a complex or mixed message but cannot be broken down into independent components. This requires more complex metaphors.

In some cases, we can obtain the more complex metaphor we require simply by extending the metaphor of the witness who may not be reliable. Shafer and Tversky [101], for example, give a metaphor for consonant belief functions [belief functions such that $m(A) = 0$ except for a nested family of subsets A] by imagining a witness whose statements may be meaningful to different degrees of exactitude, with different probabilities.

In Ref. 86 I suggested a more abstract but very general metaphor—the metaphor of a randomly coded message. In this metaphor, we have probabilities for several codes that might be used to encode a message. We do not yet know what the message says, but we know it is true. We have this message in hand in its coded form, and we will try to decode it using each code, but the probabilities are judgments we make before this decoding. When we do decode using the different codes, we sometimes get nonsense and we sometimes get a comprehensible statement. It seems sensible, in this situation, to condition our probabilities for the codes by eliminating the ones with which we get nonsense. The conditioned probability for each remaining code can then be associated with the statement we get by decoding using that code. These statements may be related in various ways; some may be inconsistent with each other, and some may be stronger than others. Thus we obtain the complexity of an arbitrary belief function.

In this metaphor, the independence of two belief functions means that two different people independently choose codes with which to send two possibly different (though both true) messages. Our uncertainties about the codes in the two cases remain independent unless possible codes imply contradictory messages. If s_1 is a possible code for the first person, and s_2 is a possible code for the second person, and the first message as decoded by s_1 contradicts the second message as decoded by s_2, then it cannot be true that these were the two codes used. We eliminate such pairs of codes and normalize the probabilities of the remaining possible pairs. The probability of each pair is then associated with the conjunction of the two implied messages. This is Dempster's rule.

Both the metaphor of the witness and the more general metaphor of the randomly coded message can be presented in a way that forestalls the interpretation of belief function degrees of belief in terms of bounds on probabilities. There is no probability model for the choice of the true message sent. The probabilities are only for the choice of codes. We might visualize these probabilities in terms of a repetition of the choice of codes, but since the true message can vary arbitrarily over this population of repetitions, the idea of this population does not lead to the idea of a true unknown probability for the true message or for the true answer to the question of interest.

An insistence on imposing an overall Bayesian model on both the truth of the message and the choice of a code will lead, of course, to the conclusion that Dempster's rule is wrong. This has been shown in detail by Williams [131] and Good [39] (see also Shafer [90]). Related attempts to relate Dempster's rule to overall Bayesian models include those of Freeling and Sahlin [36] and Baron [6].

Laskey [59] has given some examples that further illustrate the difference between belief function thinking and a Bayesian approach to the randomly coded message metaphor. In Laskey's examples, strikingly different randomly coded messages produce the same belief function. These randomly coded messages would produce quite different Bayesian probabilities if combined with Bayesian priors about the true message.

What other possibilities are there for a semantics for belief functions? Krantz [55] explored the possibility of justifying the rules for belief functions, including Dempster's rule, using measurement-type axioms. This offers at least a partial nonprobabilistic semantics for belief functions, because the axioms are concerned with comparisons of evidence and hence provide guidance in the assessment of evidential strength. I have been unable to convince myself, however, that this approach provides an understanding of the combination of evidence. In particular, it is difficult to justify Dempster's rule fully without a probabilistic basis for the concept of independence.

Pearl [74] developed a metaphor involving random switches to provide a semantics for belief functions. Again, however, I have not been able to convince myself that this metaphor provides a full basis for Dempster's rule. It does not seem to provide a basis for normalization.

In this section, I have used the word "semantics" in a process-oriented way. Semantics for a mathematical theory of evidence is guidance in using the theory to make quantitative judgments of the strength of evidence. The meaning of the resulting judgments is not independent of this process of judgment.

Classical logic and the classical frequentist and Bayesian theories of probability all have a stronger conception of semantics. In all three cases, we can say what statements of the theory mean without reference to how they are obtained. In logic, the meaning of statements can be explained in terms of possible models or worlds. In frequentist probability, the meaning of statements can be explained in terms of actual frequencies. In Bayesian probability, the meaning of statements can be explained in terms of a person's willingness to bet. I do not believe, however, that such process-independent semantics is a reasonable goal for AI. I agree with Winograd [133] that AI must use statements that have no meaning "in any semantic system that fails to deal explicitly with the reasoning process." One fundamental aspect of the subjectivity of judgments under uncertainty is the fact that these judgements depend on the process by which they are made as well as on the objective nature of the evidence.

8. SORTING EVIDENCE INTO INDEPENDENT ITEMS

Dempster's rule should be used to combine belief functions that represent independent items of evidence. But when are items of evidence independent?

How can we tell? These are probably the questions asked most frequently about belief functions.

The independence required by Dempster's rule is simply probabilistic independence applied to the questions for which we have probabilities rather than directly to the question of interest. In the metaphor of the randomly coded messages, this means that the codes are selected independently. In the more specialized metaphor of independent witnesses, it means that the witnesses (or at least their current properties as witnesses) are selected independently from well-defined populations.

Whether two items of evidence are independent in a real problem is a subjective judgment, in the belief function approach as in the Bayesian approach. There is no objective test.

In practice, our task is to sort out the uncertainties in our evidence. When items of evidence are not subjectively independent, we can generally identify what uncertainties they have in common, thus arriving at a larger collection of items of evidence that are subjectively independent. Typically, this maneuver has a cost—it forces us to refine, or make more detailed, the frame over which our belief functions are defined. We can illustrate this by adapting an example from Pearl [74].

Suppose my neighbor Mr. Watson calls me at my office to say he has heard my burglar alarm. In order to assess this testimony in belief function terms, I assess probabilities for the frame

$$S_1 = \{\text{Watson is reliable, Watson is not reliable}\}$$

Here Watson being reliable means he is honest and he can tell whether he is hearing my burglar alarm. I can use these probabilities to get degrees of belief for the frame

$$T = \{\text{My alarm sounded, My alarm did not sound}\}$$

Putting a probability of 90%, say, on Watson being reliable, I get a 90% degree of belief that my burglar alarm sounded and a 0% degree of belief that my burglar alarm did not sound.

I now call another neighbor, Mrs. Gibbons, who verifies that my alarm sounded. I can assess her testimony in the same way, by assessing probabilities for the frame

$$S_2 = \{\text{Gibbons is reliable, Gibbons is not reliable}\}$$

Suppose I also put a probability of 95% on Gibbons being reliable, so that I again obtain a 95% degree of belief that my burglar alarm sounded and a 0% degree of belief that it did not sound.

Were I to combine these two belief functions by Dempster's rule, I would obtain an overall degree of belief of 99.5% that my burglar alarm sounded. This is inappropriate, however, for the two items of evidence involve a common uncertainty—whether there might have been some other noise similar to my burglar alarm.

To deal with this problem, I must pull my skepticism about the possibility of a similar noise out of my assessment of Watson's and Gibbons's reliability, and identify my grounds for this skepticism as a separate item of evidence. So I now have three items of evidence—my evidence for Watson's honesty (I say honesty now instead of reliability, since I am not including here the judgment that there are no other potential noises in the neighborhood that Watson might confuse with my burglar alarm), my evidence for Gibbons's honesty, and my evidence that there are no potential noises in the neighborhood that sound like my burglar alarm.

These three items of evidence are now independent, but their combination involves more than the frame T. In its place, we need the frame $U = \{u_1, u_2, u_3\}$, where

$$u_1 = \text{My alarm sounded}$$

$$u_2 = \text{There was a similar noise}$$

$$u_3 = \text{There was no noise}$$

(Let us exclude, for simplicity of exposition, the possibility that there were two noises, my alarm and also a similar noise.) My first two items of evidence (my evidence for Watson's and Gibbons's honesty) both provide a high degree of belief in $\{u_1, u_2\}$, and the third item (my evidence against the existence of other noise sources) provides a high degree of belief in $\{u_1, u_3\}$. Combining the three by Dempster's rule produces a high degree of belief in $\{u_1\}$.

A Bayesian approach to this problem would be somewhat different, but it too would involve refining the frame T to U or something similar. In the Bayesian case, we would ask whether the events "Watson says he heard a burglar alarm" and "Gibbons says she heard a burglar alarm" are subjectively independent. They are not unconditionally independent, but they are independent conditional on a specification of what noise actually occurred. I can exploit this conditional independence in assessing my subjective probabilities, but in order to do so I must bring the possibility of other noise into the frame.

In the belief function approach, one talks not about conditional independence of propositions, but rather about the overlapping and interaction of evidence. For further explanation and more examples, see Shafer [82, Chap. 8; 86, 89, 94] and Srivastava et al. [113].

9. STATISTICAL INFERENCE AND FREQUENCY REASONING

Statistical inference is inference about frequencies from sample data. There are belief function methods for statistical inference, as well as Bayesian and classical frequentist methods. In this section, I discuss the relevance of statistical inference, and of frequency reasoning more generally, to the general problem of subjective judgment. We can sometimes draw analogies between nonsample evidence and imaginary data and hence use Bayesian or belief function methods of statistical inference in nonstatistical problems. To do this successfully, however, we must distinguish clearly between frequencies and degrees of belief.

The problem of statistical inference has two distinctive features, two features that distinguish it from the general problem of subjective judgment. First, the population that defines the frequencies is well defined. Second, we have substantial sample data from that population. There are many reasons for wanting to know about frequencies in particular population. Sometimes we want to know about a population for its own sake. Sometimes knowledge about frequencies in a population may serve as evidence about causal mechanisms that manifest themselves in the population. Sometimes knowledge about frequencies in a population can help us make subjective judgments about a question that can be regarded as having its answer drawn randomly from the population. In the last two cases, statistical inference is only part of the undertaking, but it is an important part.

The enduring philosophical conundrum of statistical inference is the fact that the sample data do not seem to provide a complete basis for making inferences about frequencies or about parameters in statistical models (models for frequencies). We must usually supplement these data with subjective judgment or with arbitrary choices. In the Bayesian case, we must supplement them with subjective probabilities based on evidence other than the sample data. In both the classical frequentist and belief function cases, there are usually several methods of statistical inference for a given problem, and we must make an arbitrary choice among them.

Classical frequentism is largely concerned with methods of estimating frequencies (or parameters that determine frequencies) from sample data and with estimating the average error in the estimates. There are almost always competing estimators in a given problem, and the choice among these methods is not always clear-cut (Efron [32]). Bayesian methods give more complete and definite answers; they produce probability distributions for the frequencies or parameters. These probability distributions depend, however, on prior subjective opinions as well as on the sample data (Savage [80], Lindley [63]).

Dempster, in his original work on belief functions [17–23], was motivated by the desire to obtain probability judgments based only on sample data, without dependence on prior subjective opinion. His work, together with later work on belief function statistical inference (Krantz and Miyamoto [57], Shafer [82, 83, 87, 88], Tritchler and Lockwood [119], Walley [121], Wasserman [123], Weisberg [126], and West [130]), has shown that this is possible. But this work has produced a variety of belief function methods, not a single prescribed method. In Ref. 88 I argued that we should choose among these methods by considering the nature of the evidence for the statistical model. But the evidence for the statistical model is often too nebulous for this approach to be helpful.

Most problems of subjective judgment are not problems of statistical inference, because there are neither sample data nor a well-defined population. In Section 2, I pointed out that the Bayesian approach, in general, draws an analogy between our actual evidence and knowledge of frequencies in a population. But usually this is only an analogy. The population in question is purely imaginary. Moreover, the analogy is with knowledge of frequencies, not with knowledge of sample data.

In some cases, however, we can approach a problem of subjective judgement by drawing an analogy between certain items of evidence and imaginary sample data from a real or imaginary population. This will allow us to use Bayesian or belief function statistical methods even though we do not have real sample data. Suppose, for example, that we are interested in a bird named Tweety. We want to know whether Tweety flies and whether Tweety is a penguin. We decide to make judgements about this by defining a population of birds in some way and thinking of Tweety as randomly selected from this population. We have some opinions, based on fragmentary information, hunches, partial memories, and so on, about the proportion of birds in this population that fly and the proportion that are penguins. If we can assess the strength of this evidence by saying that it is equivalent to certain sample data, then we can express this strength in terms of likelihoods, and then we can combine these likelihoods with other evidence in either a Bayesian or a belief function framework (Krantz [56]). The result would be probabilities about the frequencies in the population and, derivatively, probabilities about Tweety.

It is important, if we follow this strategy of using statistical methods for nonstatistical evidence, to distinguish clearly between frequencies and degrees of belief. The frequencies, even if they are imaginary because the population is imaginary, should not be thought of as degrees of belief, because they are not taken as known. We are evaluating evidence for what these frequencies are.

I emphasize the distinction between frequencies and degrees of belief because it tends to disappear when we use the basic Bayesian analogy, without statistical methods. The basic Bayesian analogy is between our actual evidence and an ideal picture in which frequencies are known and hence equal to our degrees of belief. It is natural, when making this analogy, to say that the numbers we produce are both degrees of belief and guesses at frequencies in an imaginary population. This basic Bayesian analogy will be needed even if we are using Bayesian statistical methods, for it is needed in assessing the prior subjective probabilities. Thus in a Bayesian statistical analysis for a nonstatistical problem, we will have some "frequencies" that are degrees of belief and other "frequencies" that are unknown.

In the case of belief functions, a careless equating of frequency with degree of belief is especially dangerous. If our strategy of subjective judgment involves applying belief function methods of statistical inference to an imaginary population, we must be clear whether a given number is a guess at a frequency or a degree of belief. Combining degrees of belief by Dempster's rule may be appropriate. Combining different guesses about frequencies by Dempster's rule certainly is not.

Consider, for example, our guesses about the proportion of birds in Tweety's population that fly and the proportion that are penguins. These guesses should not be represented as belief functions over a set of statements about Tweety and then combined by Dempster's rule.

We can explain this in terms of our discussion of dependence in the preceding section. Both guesses bear on Tweety only through their accuracy as guesses about the population. This means that they have in common the uncertainty involved in choosing Tweety at random from the population. Depending on how we obtained the guesses, they may also have other uncertainties in common. Like every problem of dependence, we can deal with this problem within the belief function approach by sorting out the uncertainties and properly refining our frame. In this case we must bring the possible values for the population frequencies into the frame. We can then formalize the connection between these frequencies and Tweety as one of our items of evidence. We must also identify our sources of evidence about the frequencies, sort out their uncertainties, and use them to assess belief functions about what these frequencies are.

Judea Pearl, in his article in this issue, gives some examples of misleading results that can arise from representing conditional probabilities as belief functions and then combining these conditional probabilities by Dempster's rule. The error involved in these examples is the same as the error in the Tweety example; we are using Dempster's rule to combine fragments of information about frequencies in a given population. A legitimate belief function treatment must deal explicitly with the unknown overall frequency distribution and with the evidence about it.

I do not believe, however, that applying belief function statistical methods to nonstatistical problems is the most promising direction for the use of belief functions. It is easy to talk about "the population of birds like Tweety." It is easy to talk about a population of repetitions for any particular observations whose evidential strength we want to assess. It is also easy to talk about causal models and conditional probabilities and populations of repetitions associated with them. But it is very hard to go beyond this talk and define these populations, even conceptually (Shafer [87]). The usually do not provide a good starting point either for Bayesian or belief function analyses.

I also do not think that purely statistical problems are the most important domain of application for belief functions. When adequate frequency information is available, there are usually many statistical methods from which to choose, and belief function methods, though reasonable, may be more complicated and unfamiliar than standard methods. Belief function representations of statistical evidence can be useful, however, when it is necessary to combine statistical and nonstatistical evidence (Dempster [26]).

Belief functions are most useful precisely when it is not sensible to try to embed the joint occurrence of all our evidence in an imaginary population of repetitions. Belief functions are appropriate when different populations of repetitions, real or imagined, justify probability judgments for different items of evidence and these populations bear on the question of interest and interact with each other in ways unique to each application. We saw this in the example of Section 4. The reliability of Sally and the reliability of Betty were both described in terms of imagined repetitions, repetitions in which Betty and Sally say many things about many things. But these reliabilities were linked to my car through what Betty and Sally said this particular time. Dempster's rule for combining their testimony was based on this unique aspect of the testimony, and it cannot be related to a population of repetitions.

10. COMPUTATION

The use of belief functions can involve challenging computational problems. In this section, I explain why belief functions are computationally complex and briefly review some ways of dealing with this complexity. These include Barnett's algorithm for the special case of belief functions focused on singletons, Thoma's fast Möbius transform, and tree propagation.

Why are belief functions computationally complex? A belief function Bel on a finite frame T with n elements is potentially far more complex than a probability measure P on the same frame. To specify a probability measure P, we need only n numbers—the probabilities $P(t)$ for each t in T. To specify a belief function Bel, we may need as many as $2^n - 2$ numbers—the degrees of belief $\mathrm{Bel}(A)$ for every proper nonempty subset of T.

In practice, belief functions based on individual items of evidence do not approach this potential degree of complexity. Even if T is very large, we will try to sort our evidence into relatively simple items, each of which says something simple about T. We may have a problem, however, when we try to combine these belief functions by Dempster's rule.

The potential problems in implementing Dempster's rule can be seen using any of the mathematical formulations of Section 5. If we work with the basic probability assignment m, for example, then the straightforward approach is to implement formulas (3), (4), and (5). Each of these formulas is exponentially complex in itself, and each of them must be applied exponentially often—once for each subset of T.

One simple way to see the complexity that results from Dempster's rule is to count the potential focal elements of the belief functions involved. A *focal element* for a belief function with basic probability assignment m is a subset A such that $m(A) \neq 0$. We might begin with belief functions with only a few focal elements. But we can see from (5) that the result of combining a belief function with r focal elements and a belief function with s focal elements will be a belief function with as many as *rs* focal elements. Thus we may obtain a very complex belief function by combining a moderate number of belief functions that individually are not complex at all.

The belief functions we want to combine are usually initially defined on many different frames. This is because the different items of evidence we consider all bear on slightly different questions. In order to combine the belief functions, however, we must consider all these questions together; we must work in a frame whose elements answer all the questions at once. This frame may be enormous, even though the initial frames are all small. It is in this enormous frame, in theory, that Dempster's rule must be carried out. This is why the computational complexity of belief functions is a real problem.

Kong [52, 54] introduced the idea of discussing belief function combination in terms of variables. Conceptually, we can think of the elements of a frame T as the possible values of a variable X. In a problem of subjective judgment, we may be primarily interested in a variable X_1, with a relatively small frame T_1. But as we saw in Section 8, the different items of evidence that we are interested in, and their interaction, may force us to consider further variables. In general, we may be forced to consider a whole list of variables, say $X_2, X_3, \ldots, X_n$, with frames $T_2, T_3, \ldots, T_n$, respectively. Perhaps each item of evidence we consider is relevant to a few of these variables and hence each belief function we assess involves only the frame for a few variables. The first item of evidence might be relevant to the variables X_1, X_4, and X_5, for example, and hence be representable by a belief function on the Cartesian product $T_1 \times T_4 \times T_5$. But if we want to combine all the evidence, we find ourselves working in the frame $T_1 \times T_2 \times T_3 \times \cdots \times T_n$, which may be enormous if n is large.

Barnett's Algorithm

Barnett [5] gave algorithms for Dempster's rule for the special case in which each belief function supports either a single element of T or the complement of a single element. (More precisely, the focal elements for each belief function are either singletons or complements of singletons.) These algorithms are much better than the general algorithms; they are linear rather than exponential in the frame size.

In order to understand the potential and the limits of Barnett's algorithm, we must take into account the multivariate nature of belief function computation. Individual items of evidence may in some cases produce belief functions that satisfy Barnett's conditions on a given frame. But as we study our problem, we will bring in new variables with new frames. A belief function that has singletons and their complements as focal elements in the frame $T_1 \times T_4 \times T_5$ will not have singletons and their complements as focal elements in the more refined frame $T_1 \times T_2 \times T_3 \times \cdots \times T_n$. In this more refined frame, the focal elements will be elements of a given partition and their complements. Thinking of the belief function in this larger frame does not invalidate Barnett's algorithm; it can be thought of as working on the partition. But a different item of evidence, assessed as a belief function satisfying Barnett's conditions on $T_3 \times T_4 \times T_6$, say, will involve a different partition in $T_1 \times T_2 \times T_3 \times \cdots \times T_n$, and Barnett's algorithm will not enable us to combine the two belief functions.

In practice, therefore, Barnett's algorithm must be supplemented by techniques for combining belief functions on related frames. Shafer and Logan [97] successfully used Barnett's algorithm in this way in the case of hierarchical evidence.

Fast Möbius Transforms

With care, we can reduce the complexity of Dempster's rule even in the general case, because much of the computation in applying (3), (4), and (5) for each subset is repetitive. Thoma [118] has shown how this repetition can be eliminated; the result is a fast Möbius transform, analogous to the fast Fourier transform.

It remains to be seen how useful the fast Möbius transform will be in practice. It is clear, however, that it is not enough to make arbitrary belief function computations feasible. Even reducing the computational complexity from exponential to linear is not enough if the frame is enormous.

Propagation in Trees

More recent work on computation has focused on exploiting the pattern of evidence to reduce computation on very large frames to computation on many smaller frames. This can be done whenever individual items of evidence bear directly only on clusters of variables, and these clusters can be arranged in a join tree (Shafer et al. [102], Dempster and Kong [28], Shafer and Shenoy [98]).

A tree in which the nodes are clusters of variables is called a *join tree* if whenever a variable is contained in two nodes of the tree it is also contained in all the nodes on the path between these two nodes. Figure 1 shows an example of such a tree.

Suppose we have independent items of evidence bearing on each cluster of variables in this tree. Each of these we represent by a belief function on the corresponding frame. The evidence bearing on X_2, X_3, and X_4, for example, is represented by a belief function on $T_2 \times T_3 \times T_4$. We want to combine all these belief functions on $T_1 \times T_2 \times T_3 \times \cdots \times T_{10}$. It turns out that we can do this (at least we can get the values of the resultant belief function for statements about individual variables and the given clusters of variables) by applying Dempster's rule repeatedly within the smaller frames corresponding to the clusters in the tree. The information from one cluster relevant to another cluster can be passed along the path between the two by means of messages between neighboring

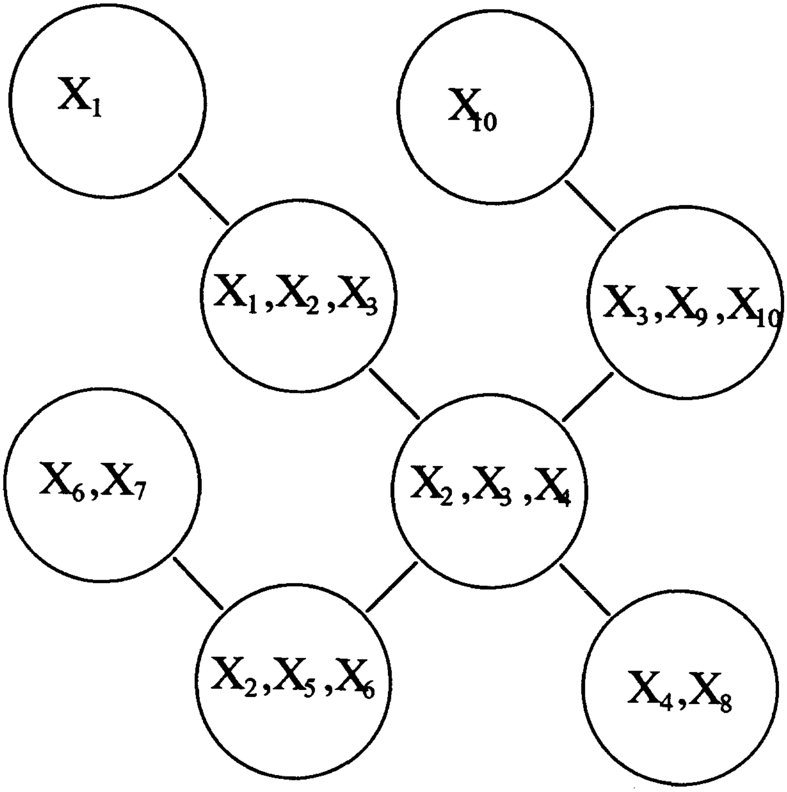

Figure 1. A join tree with 10 variables.

nodes that take the form of belief functions for the variables that the neighboring nodes have in common.

Work on belief functions in trees was initiated by the work of Gordon and Shortliffe [41, 42]. They were concerned with the problem of combining evidence by Dempster's rule when different items of evidence are relevant to different levels of specificity in a hierarchy of diseases. Gordon and Shortliffe's suggestion of approximating Dempster's rule in this case was strengthened by Shafer and Logan [97] to an efficient method of computing the exact results of the rule. Shenoy and Shafer [106], Kong [52], and Shafer et al. [102] explained how this treatment of hierarchical evidence can be understood as a special case of propagation in trees.

Bayesian and Belief Function Propagation

Interest in belief function propagation was inspired not only by Gordon and Shortliffe's work but also by Pearl's work on Bayesian propagation [73], which was based on models of conditional independence. How is the algorithm for belief function propagation in trees related to algorithms for Bayesian propagation in trees developed by Pearl and by Lauritzen and Spiegelhalter [61]?

It is possible to regard the Bayesian algorithms as special cases of the belief function algorithm (Shenoy and Shafer [106], Shafer et al. [102], Zarley et al. [138]). In order to do this, however, each probability transition matrix, which provides the probabilities for one node as conditional probabilities given another node, must be represented by a belief function. This can be done, but the way in which it is done is neither efficient nor conceptually satisfying. The probability transition matrices represent different fragments of information about a single frequency distribution, and, as I explained in Section 9, such fragments should not, in general, be represented by belief functions and combined by Dempster's rule.

Rather than treating the Bayesian algorithm as a special case of the belief function algorithm, therefore, we should think of them as two special cases of a more general algorithm. The scope of this more general algorithm has been demonstrated by Shenoy and Shafer [107]. It applies in any case where we have operations of marginalization and combination that satisfy a few simple axioms (Shenoy [104, 105]). As it turns out, this viewpoint clarifies Bayesian as well as belief function propagation (Shafer and Shenoy [98, 99]).

In both the belief function and Bayesian cases, join trees are more than computational tools. They are also conceptual tools, tools that we use in sorting out our evidence. In the Bayesian case, they provide a graphical representation of the conditional independence structure that is needed to make probability judgments manageable. In the belief function case, they provide a representation of the sorting into independent uncertainties discussed in Section 8 (Dempster [26]).

Networks

When we sort our evidence into independent items, the clusters of variables that result may be such that they cannot be arranged in a join tree. This would be case in Figure 1, for example, if the variable X_3 were added to the cluster X_6, X_7. How to handle the computational problem in this case is a difficult problem in general.

The most satisfactory solution to the problem, if possible, is to find a join tree with slightly larger clusters such that each of the clusters with which we began can fit into a cluster in the tree. In the case of Figure 1 with X_3 added to the cluster X_6, X_7, for example, we can obtain such a tree by also adding X_3 to X_2, X_5, X_6.

Bayesian propagation also sometimes requires that we embed clusters of variables in a join tree. Lauritzen and Spiegelhalter [61] give one way of doing this; Kong [52, 53] gives another way, which usually results in a more manageable tree. How to embed collections of clusters in trees in the most efficient way is the subject of a growing literature (Rose [75], Bertele and Brioschi [9], Tarjan and Yannakakis [117], Arnborg et al. [4], Mellouli [68], Zhang [140]).

11. IMPLEMENTING BELIEF FUNCTIONS IN ARTIFICIAL INTELLIGENCE

Belief functions have been implemented in a wide variety of expert systems, and I am not prepared to evaluate or even list these implementations. The implementations I find most convincing, however, are those designed for interactive use. These include Gister (Lowrance et al. [65]), Russell Almond's program (Almond [1]), DELIEF (Zarley et al. [13]), AUDITOR's ASSISTANT (Shafer et al. [103]), and MacEvidence (Hsia and Shenoy [44]). These systems help human users build and evaluate belief networks. They require the user to make the judgments of independence that justify the network and to provide the numerical judgments of support based on each item of evidence.

Interactive systems seem appropriate to belief functions, since the theory practically requires that the relation between evidence and questions of interest should be unique to each application. Many probabilistic systems—such as the HUGIN system for medical diagnosis (Anderson et al. [2])—apply the same conditional independence structure and, for the most part, the same numerical judgments to each new case. This means relating the entire structure of the evidence in each case to the same population of repetitions. As I have already argued, belief functions are appropriate to situations where this direct frequentist application of probability is not possible—situations where different populations of repetitions, real or imagined, justify probability judgments for different items of evidence, and these populations bear on the question of interest and interact with each other in ways unique to each application. Such uniqueness means that the belief network and the numerical judgments must be constructed anew for each case.

The next step, of course, is to automate this constructive process. This is difficult, but some progress has been made (Lowrance [64], Andress and Kak [3], Biswas and Anand [10], Laskey et al. [60]).

Ultimately, it would be desirable to automate not only the construction of belief networks but also numerical assessment. I am not aware of current work in this direction, but current ideas in distributed memory (Kanerva [48]) encourage the idea that there is enough independence in such memories to permit the use of Dempster's rule.

12. OTHER TOPICS

In this final section, I briefly discuss several topics: generalizations of belief function theory, decision methods based on belief functions, methods for reaching consensus using belief functions, work on belief function weights of evidence, and other mathematical work, especially on infinite frames.

Generalizations

One way to generalize belief function theory is to retain the class of belief functions and generalize Dempster's rule for combining belief functions. I have discussed some generalizations that allow for dependent evidence [93, 94], but I am not aware of practical applications of these generalizations. In my view, the theory of belief functions should be used as a way of examining evidence (Dempster [25]), and in the examples I have thought about, this seems to lead to sorting out the independencies in the evidence, as in the example of Section 8.

Another approach is to retain the class of belief functions but generalize the idea of a compatibility relation to allow for partial or probabilistic compatibility (Kohlas [51], Yen [135]). In some cases, this idea can be seen not as a generalization at all, but rather as an introduction of frames intermediate between the S and T of Section 5. In other cases, the probabilistic basis of belief functions is lost, and the rationale for the generalization is not clear (Dubois and Prade [29]).

Finally, we can generalize the class of belief functions to more general set functions. Going in one direction, this takes us to the lower-probability theory discussed in Section 6 and even to versions of lower-probability theory that generalize probability bounds (Walley and Fine [122], Wasserman and Kadane [125]). Going in another direction, away from probability but toward a variety of rules of combination, we enter the vast literature on fuzzy sets (Dubois and Prade [29, 30]).

Decision Methods

Tom Strat, in his article in this volume, sets out a natural approach to decision theory within the belief function framework. The basic idea, that of upper and lower expectations, can also be found in Choquet [13] and Dempster [17]. Examples and practical elaborations are given by Wesley [127, 128] and Wesley et al. [129].

Variations on this basic idea are possible, but I am not aware of extensive explorations of them. Two brief explorations can be found in Shafer [91, response to discussion] and Dempster and Kong [27].

Another approach to relating belief functions to decision theory is to apply the usual von Neumann–Morgenstern theory to the linear space of belief functions. Jaffray [46] has developed this approach.

Consensus for Belief Functions

Rationales for average or consensus belief functions for group decision have been developed by Wagner [120] and Williams [132]. In Ref. 92, however, I argued against global consensus methods and in favor of direct reassessment by the group of the different items of evidence that contribute to a belief function.

Infinite Frames

Because my 1976 book dealt only with the case where the frame T is finite, some readers were left with the impression that the theory of belief functions had been developed only for this case. In fact, however, Dempster's articles had dealt from the outset with the continuous case. This is natural in the context of parametric statistical inference, since even the set of possible values for a single unknown probability is a continuous interval.

Dempster's early work included a treatment of the mathematics of belief functions generated by random closed intervals (Dempster [20]). This topic was also treated by Strat [114].

In general, the mathematical study of belief functions on continuous spaces will employ various regularity conditions, analogous to countable additivity or to familiar topological conditions in probability theory. These include topological conditions (Zhang and Wang [114]) or mere sequential continuity, analogous to countable additivity (Ross [76]). My own contributions to belief functions [83, 85] on continuous frames emphasized the condition of condensability, which was satisfied by the belief functions studied by Dempster. A condensable belief function is one for which the plausibility of every set A can be approximated by *finite* subsets of A; this expresses very strongly the intuition that in the case of subjective judgment, continuous mathematics should not be considered more than a convenient approximation to something fundamentally finite.

Weights of Evidence

My original purpose in writing the 1976 monograph was to explain an unsolved mathematical problem, which I called the *weight of conflict conjecture*. This conjecture derived from the idea of associating a weight of evidence to a belief function that supports a single subset of a frame to a certain degree. Such a belief function is called a *simple support function*. If A is the subset it supports, then the weight of evidence in favor of A is $-\log[1 - \mathrm{Bel}(A)]$. Weights of evidence add when simple support functions are combined by Dempster's rule. As I showed, all belief functions on finite frames can be obtained by combining simple support functions, reducing to coarser frames, and taking limits. Thus a belief function can always be understood in terms of a collection of weights of evidence. There is a possibility of competing representations by weights of evidence, however, because a given belief function can be obtained by coarsening in more than one way. The weight-of-conflict conjecture, if true, pointed to a way in which the competing representations would be in agreement. This conjecture remains unsettled, but the conjectured agreement was established by a different argument by Zhang [139]. The possibility of using weights of evidence in belief function applications remains largely unexploited.

Other Mathematical Advances

Berres [7] showed that products of belief functions are always belief functions. This has possible applications in the context of discounting belief functions (Shafer [87]). Berres [8] gave a comprehensive review of the relation between belief functions and Sugeno's λ-additive measures (Sugeno [115]). Another interesting advance is the extension to belief functions of ideas of information and entropy that have been associated with probability distributions (Smets [108], Yager [134]).

ACKNOWLEDGMENTS

Research for this article was partially supported by the National Science Foundation through grant IRI8902444. In preparing this article, I have profited from conversations with Paul Cohen, Pierre Ndilikilikesha, Judea Pearl, Prakash Shenoy, and Lianwen Zhang.

References

1. Almond, R., Fusion and propagation of graphical belief models, Doctoral Dissertation, Research Report S-130, Dept. of Statistics, Harvard University, 1989.
2. Anderson, S. K., Olesen, K. G., Jensen, F. V., and Jensen, F., HUGIN—a shell for building Bayesian belief universes for expert systems, *IJCAI-89*, 1080–1085, 1989.
3. Andress, K., and Kak, A., Evidence accumulation and flow of control, *AI Mag.* 9(2), 75–94, 1988.
4. Arnborg, S., Corneil, D. G., and Proskurowski, A., Complexity of finding embeddings in a k-tree, *SIAM J. Algebraic Discrete Methods* 8, 277–284, 1987.
5. Barnett, J. A., Computational methods for a mathematical theory of evidence, *IJCAI-81*, 868–875, 1981.
6. Baron, J., Second-order probabilities and belief functions, *Theory Decision* 23, 25–36, 1987.

7. Berres, M., On a multiplication and a theory of integration for belief and plausibility functions, *J. Math. Anal. Appl.* 121, 487–505, 1987.

8. Berres, M., λ-Additive measures on measure spaces, *Fuzzy Sets Syst.* 27, 159–169, 1988.

9. Bertele, U., and Brioschi, F., *Nonserial Dynamic Programming*, Academic, New York, 1972.

10. Biswas, G., and Anand, T. S., Using the Dempster-Shafer scheme in a mixed-initiative expert system shell, in *Uncertainty in Artificial Intelligence*, Vol. 3 (L. N. Kanal, T. S. Levitt, and J. F. Lemmer, Eds.), North-Holland, New York, 1989, pp. 223–239.

11. Brooks, R. A., A robust layered control system for a mobile robot, *IEEE J. Robot. Autom.* 2(1), 14–23, 1986.

12. Cheeseman, P., An inquiry into computer understanding (with discussion), *Comput. Intell.* 4, 58–66, 1988.

13. Choquet, G., Theory of capacities, *Ann. Inst. Fourier, Grenoble* 5, 131–295, 1953.

14. Cohen, P. R., The control of reasoning under uncertainty: a discussion of some programs, *Knowledge Eng. Rev.* 2, 5–25, 1987.

15. Cox, R. T., *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, 1961.

16. Dawes, R. M., Faust, D., and Meehl, P. E., Clinical versus actuarial judgment, *Science* 243, 1668–1674, 1989.

17. Dempster, A. P., New methods for reasoning towards posterior distributions based on sample data, *Ann. of Math. Stat.* 37, 355–374, 1966.

18. Dempster, A. P., Upper and lower probabilities induced by a multivalued mapping, *Ann. Math. Stat.* 38, 325–339, 1967.

19. Dempster, A. P., Upper and lower probability inferences based on a sample from a finite univariate population, *Biometrika* 54, 515–528, 1967.

20. Dempster, A. P., Upper and lower probabilities generated by a random closed interval, *Ann. Math. Stat.* 39, 957–966, 1968.

21. Dempster, A. P., A generalization of Bayesian inference (with discussion), *J. Roy. Stat. Soc. Ser. B* 30, 205–247, 1968.

22. Dempster, A. P., Upper and lower probability inferences for families of hypotheses with monotone density ratios, *Ann. Math. Stat.* 40, 953–969, 1969.

23. Dempster, A. P., A class of random convex polytopes, *Ann. Math. Stat.* 43, 260–272, 1972.

24. Dempster, A. P., Comment, *J. Am. Stat. Assoc.* 77, 339–341, 1982.

25. Dempster, A. P., Probability, evidence, and judgment (with discussion), in *Decision Making: Descriptive, Normative, and Prescriptive Interactions* (D. E.

Bell, H. Raiffa, and A. Tversky, Eds.), Cambridge Univ. Press, New York, 1988, pp. 284–298.

26. Dempster, A. P., Construction and local computation aspects of network belief functions, in *Influence Diagrams, Belief Nets, and Decision* Analysis (R. M. Oliver and J. Q. Smith, Eds.), Wiley, New York, 1990, Chap. 6.

27. Dempster, A. P., and Kong, A., Discussion of papers by Lindley, Shafer, and Spiegelhalter, *Stat. Sci.* 1, 32–36, 1987.

28. Dempster, A. P., and Kong, A., Uncertain evidence and artificial analysis, *J. Stat. Planning Inference* 20, 355–368, 1988.

29. Dubois, D., and Prade, H., A set-theoretic view of belief functions, *Int. J. Gen. Syst.* 12, 193–226, 1986.

30. Dubois, D., and Prade, H., *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum, New York, 1988.

31. Duda, R. O., and Hart, P. E., *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

32. Efron, B., Controversies in the foundations of statistics, *Am. Math. Monthly* 85, 231–246, 1978.

33. Fagin, R., and Halpern, J. Y., Uncertainty, belief, and probability, *IJCAI-89*, 1161–1167, 1989.

34. Fagin, R., and Halpern, J. Y., Updating beliefs vs. combining beliefs, unpublished report, 1989.

35. Fishburn, P. C., Analysis of decisions with incomplete knowledge of probabilities, *Oper. Res.* 13, 217–237, 1965.

36. Freeling, A. N. S., and Sahlin, N.-E., Combining evidence, in *Evidentiary Value* (P. Gärdenfors, B. Hansson, and N.-E. Sahlin, Eds.), Gleerups, Lund, Sweden, 1983.

37. Gärdenfors, P., Hansson, B., and Sahlin, N.-E., Eds., *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, Gleerups, Lund, Sweden, 1983.

38. Good, I. J., The measure of a non-measurable set, in *Logic, Methodology, and Philosophy of Science* (Nagel, Suppes, and Tarski, Eds.), Stanford Univ. Press, Stanford, Cal., 1962, pp. 319–326.

39. Good, I. J., Discussion of "Lindley's paradox," by Glenn Shafer, *J. Am. Stat. Assoc.* 77, 342–344, 1982.

40. Goodman, I. R., and Nguyen, H. T., *Uncertainty Models for Knowledge-Based Systems*, North-Holland, New York, 1985.

41. Gordon, J., and Shortliffe, E. H., The Dempster-Shafer theory of evidence, in *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project* (B. G. Buchanan and E. H. Shortliffe, Eds.), Addison-Wesley, Reading, Mass., 1984.

42. Gordon, J., and Shortliffe, E. H., A method for managing evidential reasoning in hierarchical hypothesis spaces, *AI* 26, 323-358, 1985.

43. Halmos, P. R., *Measure Theory*, Van Nostrand, Princeton, N.J., 1950.

44. Hsia, Y.-T., and Shenoy, P. P., An evidential language for expert systems, in *Methodologies for Intelligent Systems*, Vol. 4 (Z. Ras, Ed.), North-Holland, New York, 1989, pp. 9-16.

45. Isaacs, H. H., Sensitivity of decisions to probability estimation errors, *Oper. Res.* 11, 536-551, 1963.

46. Jaffray, J.-Y., Linear utility theory for belief functions, *Oper. Res. Lett.* 8, 107-112, 1989.

47. Kadane, J., *Robustness in Bayesian Statistics*, North-Holland, New York, 1984.

48. Kanerva, P., *Sparse Distributed Memory*, MIT Press, Cambridge, Mass., 1988.

49. Kelley, J., *General Topology*, Van Nostrand-Reinhold, New York, 1955.

50. Kofler, E., Kmietowicz, Z. W., and Pearman, A. D., Decision making with linear partial information, *J. Oper. Res. Soc.* 35, 1079-1090, 1984.

51. Kohlas, J., Conditional belief structures, Tech. Report 131, University of Fribourg, Switzerland, 1987.

52. Kong, A., Multivariate belief functions and graphical models, Doctoral dissertation, Dept. of Statistics, Harvard Univ., 1986.

53. Kong, A., Construction of a tree of cliques from a triangulated graph, Tech. Report S-118, Dept. of Statistics, Harvard Univ., 1986.

54. Kong, A., A belief function generalization of Gibbs ensembles, Tech. Report 239, Dept. of Statistics, Univ. of Chicago, 1988.

55. Krantz, D., Foundations of the theory of evidence, Invited address, Society for Mathematical Psychology, 1982.

56. Krantz, D., Judgments of likelihood evidence, Tech. Report, Dept. of Psychology, Columbia Univ., 1990.

57. Krantz, D., and Miyamoto, J., Priors and likelihood ratios as evidence, *J. Am. Stat. Assoc.* 78, 418-423, 1983.

58. Kyburg, H. E., Bayesian and non-Bayesian evidential updating, *AI* 31, 297-294, 1987.

59. Laskey, K. B., Belief in belief functions: an examination of Shafer's canonical examples, in *Uncertainty in Artificial Intelligence*, Vol. 3 (L. N. Kanal, T. S. Levitt, and J. F. Lemmer, Eds.), North-Holland, New York, 1989.

60. Laskey, K. B., Cohen, M. S., and Martin, A. W., Representing and eliciting knowledge about uncertain evidence and its implications, *IEEE Trans. Syst., Man, Cybern.* 19, 536-557, 1989.

61. Lauritzen, S. L., and Spiegelhalter, D. J., Local computations with probabilities

on graphical structures and their application to expert systems (with discussion), *J. Roy. Stat. Soc. Ser. B* 50, 157–224, 1988.

62. Levi, I., *The Enterprise of Knowledge*, MIT Press, Cambridge, Mass., 1980.
63. Lindley, D. V., *Bayesian Statistics: A Review*, SIAM, Philadelphia, 1972.
64. Lowrance, J. D., Automated argument construction, *J. Stat. Planning Inference* 20, 369–387, 1988.
65. Lowrance, J. D., Garvey, T. D., and Strat, T. M., A framework for evidential-reasoning systems, *AAAI-86*, 896–903, 1986.
66. McDermott, D. V., Critique of pure reason (with discussion), *Comput. Intell.* 3, 151–160, 1987.
67. Matheron, G., *Random Sets and Integral Geometry*, Wiley, New York, 1975.
68. Mellouli, K., On the propagation of beliefs in networks using the Dempster-Shafer theory of evidence, Doctoral dissertation, School of Business, Univ. of Kansas, 1987.
69. Neveu, J., *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco, 1965.
70. Nguyen, H. T., On random sets and belief functions, *J. Math. Anal. Appl.* 65, 531–542, 1978.
71. Nilsson, N. J., *Learning Machines*, McGraw-Hill, New York, 1965.
72. Nilsson, N. J., Probabilistic logic, *AI* 28, 71–87, 1986.
73. Pearl, J., Fusion, propagation and structuring in belief networks, *AI* 29, 241–288, 1986.
74. Pearl, J., *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo, Cal., 1988.
75. Rose, D. J., Triangulated graphs and the elimination process, *J. Math. Anal. Appl.* 32, 597–609, 1970.
76. Ross, D., Random sets without separability, *Ann. Probab.* 14, 1064–1069, 1986.
77. Rota, G.-C., Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie Verw. Geb.* 2, 340–368, 1964.
78. Ruspini, E. H., Epistemic logics, probability, and the calculus of evidence, *IJCAI-87*, 924–931, 1987.
79. Savage, L. J., *The Foundations of Statistics*, Wiley, New York, 1954.
80. Savage, L. J., The foundations of statistics reconsidered, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (J. Neyman, Ed.), Univ. of California Press, 1961, pp. 575–586.
81. Shafer, G., Allocations of probability, Doctoral dissertation, Dept. of Statistics, Princeton Univ., 1973.

82. Shafer, G., *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, N.J., 1976.

83. Shafer, G., A theory of statistical evidence, in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. 2 (W. L. Harper and C. A. Hooker, Eds.), Reidel, Dordrecht, Holland, 1976, pp. 365–436.

84. Shafer, G., Non-additive probabilities in the work of Bernoulli and Lambert, *Arch. History Exact Sci.* 19, 309-370, 1978.

85. Shafer, G., Allocations of probability, *Ann. Probab.* 7, 827–839, 1979.

86. Shafer, G., Constructive probability, *Synthese* 48, 1–60, 1981.

87. Shafer, G., Lindley's paradox (with discussion), *J. Am. Stat. Assoc.* 77, 325–351, 1982.

88. Shafer, G., Belief functions and parametric models (with discussion), *J. Roy. Stat. Soc. Ser. B* 44, 322–352, 1982.

89. Shafer, G., The problem of dependent evidence, Working Paper 164, School of Business, Univ. of Kansas, 1984.

90. Shafer, G., The combination of evidence, *Int. J. Intell. Syst.* 1, 155–179, 1986.

91. Shafer, G., Savage revisited (with discussion), *Stat. Sci.* 1, 463-501, 1986.

92. Shafer, G., Discussion of "Combining probability distributions," by C. Genest and J. V. Zidek, *Stat. Sci.* 1, 135–137, 1986.

93. Shafer, G., Probability judgment in artificial intelligence and expert systems (with discussion), *Stat. Sci.* 2, 3–44, 1987.

94. Shafer, G., Belief functions and possibility measures, in *The Analysis of Fuzzy Information*, Vol. 1 (J. C. Bezdek, Ed.), CRC Press, Boca Raton, Fla., 1987.

95. Shafer, G., Discussion of "An inquiry into computer understanding," by Peter Cheeseman, *Comput. Intell.* 4, 121–124, 1988.

96. Shafer, G., The unity of probability, in *Acting Under Uncertainty: Multidisciplinary Conceptions* (G. von Furstenberg, Ed.), Kluwer, 1990, pp. 95–126.

97. Shafer, G., and Logan, R., Implementing Dempster's rule for hierarchical evidence, *AI* 33, 271–298, 1987.

98. Shafer, G., and Shenoy, P. P., Local computation in hypertrees, Working Paper 201, School of Business, Univ. of Kansas, 1988.

99. Shafer, G., and Shenoy, P. P., Probability propagation, *Ann. Math. AI*, to be published.

100. Shafer, G., and Srivastava, R. P., The Bayesian and belief-function formalisms: a general perspective for auditing, *Auditing: J. Pract. Theory* 9 (Suppl.), to be published.

101. Shafer, G., and Tversky, A., Languages and designs for probability judgment, *Cogn. Sci.* 9, 309-339, 1985.

102. Shafer, G., Shenoy, P. P., and Mellouli, K., Propagating belief functions in qualitative Markov trees, *Int. J. Approximate Reasoning* 1, 349–400, 1987.

103. Shafer, G., Shenoy, P. P., and Srivastava, R. P., AUDITOR'S ASSISTANT: a knowledge-engineering tool for audit decisions, *Proceedings of the 1988 Toche-Ross/University of Kansas Symposium*, 1988, pp. 61–84.

104. Shenoy, P. P., A valuation-based language for expert systems, *Int. J. Approximate Reasoning* 3, 383–411, 1989.

105. Shenoy, P. P., On Spohn's rule for revision of beliefs, Working Paper 213, School of Business, Univ. of Kansas, 1989.

106. Shenoy, P. P., and Shafer, G., Propagating belief functions using local computations, *IEEE Expert* 1 (3), 43–52, 1986.

107. Shenoy, P. P., and Shafer, G., Axioms for probability and belief-function propagation, in *Uncertainty in Artificial Intelligence*, Vol. 4 (R. Shachter, T. S. Levitt, J. F. Lemmer, and L. N. Kanal, Eds.), North-Holland, New York, 1990.

108. Smets, P., Information content of an evidence, *Int. J. Man-Mach. Stud.* 19, 33–43, 1983.

109. Smets, P., Belief functions, in *Non-Standard Logics for Automated Reasoning* (P. Smets, A. Mamdani, D. Dubois, and H. Prade, Eds.), Academic, New York, 1988, pp. 253–286.

110. Smith, C. A. B., Consistency in statistical inference and decision (with discussion), *J. Roy. Stat. Soc. Ser. B* 23, 1–25, 1961.

111. Snow, P., Bayesian inference without point estimates, *AAAI-86*, 233–237, 1986.

112. Speigelhalter, D. J., A statistical view of uncertainty in expert systems, in *Artificial Intelligence and Statistics* (W. Gale, Ed.), Addison-Wesley, Reading, Mass., 1986, pp. 17–56.

113. Srivastava, R. P., Shenoy, P. P., and Shafer, G., Audit risk: a belief-function approach, Working Paper 212, School of Business, Univ. of Kansas, 1989.

114. Strat, T., Continuous belief functions for evidential reasoning, *AAAI-84*, 308–313, 1984.

115. Sugeno, M., Theory of fuzzy integrals and its applications, Doctoral dissertation, Tokyo Institute of Technology, 1974.

116. Suppes, P., The measurement of belief, *J. Roy. Stat. Soc. Ser. B* 36, 160–191, 1974.

117. Tarjan, R. E., and Yannakakis, M., Simple linear time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, *SIAM J. Comput.* 13, 566–579, 1984.

118. Thoma, H. M., Factorization of belief functions, Doctoral dissertation, Dept. of Statistics, Harvard Univ., 1989.

119. Trichtler, D., and Lockwood, G., Modelling the reliability of paired comparisons, *J. Math. Psychol.*, to appear.

120. Wagner, C., Consensus for belief functions and related uncertainty measures, *Theory Decision*, to appear.

121. Walley, P., Belief-function representations of statistical evidence, *Ann. Stat.* 15, 1439–1465, 1987.

122. Walley, P., and Fine, T. L., Towards a frequentist theory of upper and lower probability, *Ann. Stat.* 10, 741–761, 1982.

123. Wasserman, L. A., Belief functions and statistical inference, Tech. Report 458, Dept. of Statistics, Carnegie Mellon Univ., 1989.

124. Wasserman, L. A., Prior envelopes based on belief functions, *Ann. Stat.*, 18, 454–464, 1990.

125. Wasserman, L. A., and Kadane, J., Bayes's theorem for Choquet capacities, *Ann. Stat.*, to appear.

126. Weisberg, H., Upper and lower probability inferences from ordered multinomial data, Doctoral dissertation, Dept. of Statistics, Harvard Univ., 1969.

127. Wesley, L. P., Reasoning about control: an evidential approach, Tech. Report 324, AI Center, SRI International, Menlo Park, Cal., 1984.

128. Wesley, L. P., Evidential knowledge-based computer vision, *Opt. Eng.*, 25(3), 363–379, 1986.

129. Wesley, L. P., Lowrance, J. D., and Garvey, T. D., Reasoning about control: the investigation of an evidential approach, *IJCAI-83*, 203–206, 1983.

130. West, S., Upper and lower probability inference for the logistic function, Doctoral dissertation, Dept. of Statistics, Harvard Univ., 1971.

131. Williams, P. M., Discussion of "Belief functions and parametric models," by Glenn Shafer, *J. Roy. Stat. Soc. Ser. B* 44, 341–342, 1982.

132. Williams, P. M., Averaging and eliciting expert opinion, *AI*, to appear.

133. Winograd, T., Extended inference modes in reasoning by computer systems, *AI* 13, 5–26, 1980.

134. Yager, R. R., Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Syst.* 9, 249–260, 1983.

135. Yen, J., GERTIS: a Dempster-Shafer approach to diagnosing hierarchical hypotheses, *Commun. ACM* 32(5), 573–585, 1989.

136. Zadeh, L. A., Fuzzy sets as a basis for a theory of possibility, *Fuzzy Sets Syst.* 1, 3–28, 1978.

137. Zadeh, L. A., A simple view of the Dempster-Shafer theory of evidence and its implications for the rule of combination, *AI Mag.* 7(2), 85–90, 1986.

138. Zarley, D., Hsia, Y.-T., and Shafer, G., Evidential reasoning using DELIEF, *AAAI-88*, 205–209, 1988.

139. Zhang, L., Weights of evidence and internal conflict for support functions, *Inf. Sci.* 38, 205–212, 1986.

140. Zhang, L., Studies on hypergraphs, Working Paper 215, School of Business, Univ. of Kansas, 1990.

141. Zhang, L., and Wang, P., A generalization of Choquet's theorem to measurable spaces, *Kexue Tongbao* 32, 1596–1601, 1987.